Data management challenges in analysis and synthesis in the ecosystem sciences

A. Specht [a,*], S. Guru [a,b], L. Houghton [a], L. Keniger [c], P. Driver [d,e], E.G. Ritchie [f], K. Lai [a], A. Treloar [g]

[a] Australian Centre for Ecological Analysis and Synthesis, a facility of the Terrestrial Ecosystem Research Network, University of Queensland, Australia
[b] Terrestrial Ecosystem Research Network, University of Queensland, Australia
[c] School of Biological Sciences, University of Queensland, Australia
[d] Office of Water, NSW Department of Primary Industries, Orange, New South Wales, Australia
[e] Centre for Ecosystem Science, University of New South Wales, Kensington, NSW, Australia
[f] Centre for Integrative Ecology, School of Life and Environmental Sciences, Deakin University, Burwood, Victoria, Australia
[g] Australian National Data Service, Monash University, Australia

HIGHLIGHTS

- Transdisciplinary collaboration is difficult, but delivers innovative products.
- Be disciplined: use an end-to-end Data Management Plan.
- Change the norms: sharing and publishing data requires a new set of skills.
- Anticipate Reuse: encourage colleagues to agree on minimum elements for dataset reusability.

ARTICLE INFO

ABSTRACT

Open-data has created an unprecedented opportunity with new challenges for ecosystem scientists. Skills in data management are essential to acquire, manage, publish, access and re-use data. These skills span many disciplines and require trans-disciplinary collaboration.

Science synthesis centres support analysis and synthesis through collaborative 'Working Groups' where domain specialists work together to synthesise existing information to provide insight into critical problems. The Australian Centre for Ecological Analysis and Synthesis (ACEAS) served a wide range of stakeholders, from scientists to policy-makers to managers. This paper investigates the level of sophistication in data management in the ecosystem science community through the lens of the ACEAS experience, and identifies the important factors required to enable us to benefit from this new data-world and produce innovative science.

ACEAS promoted the analysis and synthesis of data to solve transdisciplinary questions, and promoted the publication of the synthesised data. To do so, it provided support in many of the key skillsets required. Analysis and synthesis in multi-disciplinary and multi-organisational teams, and publishing data were new for most. Data were difficult to discover and access, and to make ready for analysis, largely due to lack of metadata. Data use and publication were hampered by concerns about data ownership and a desire for data citation. A web portal was created to visualise geospatial datasets to maximise data interpretation. By the end of the experience there was a significant increase in appreciation of the importance of a Data Management Plan.

It is extremely doubtful that the work would have occurred or data delivered without the support of the Synthesis centre, as few of the participants had the necessary networks or skills. It is argued that participation in the Centre provided an important learning opportunity, and has resulted in improved knowledge and understanding of good data management practices.

Crown Copyright © 2015 Published by Elsevier B.V. All rights reserved.

1. Introduction

The size and number of datasets are increasing at a tremendous rate due to advancement in technologies used in data collection and collation activities. This has created an opportunity to address some of the complex, multi-dimensional environmental problems, which require multi-, inter- and trans-disciplinary approaches (Peters, 2010). Policy and decision makers expect clear, understandable information on urgent environmental issues such as climate change, natural resource depletion, biodiversity loss and environmental health. Access

* Corresponding author.

to a wide variety of information comes attached with complexity, transparency, integrity and interpretation problems.

There is a wealth of data emerging in the ecosystem sciences from existing sources, and new data are being created all the time. In this domain, every observation is unique, due to its context in space and time. If the information on an observation is lost, it is lost forever because it is almost impossible to measure the observation again in the original context (Ellison, 2010). This is one of the greatest motivations for the re-use of existing data for knowledge creation. With the advancement of technology in capturing and processing data, we have reached the fourth paradigm of data-intensive science and communication, where collaboration between different domain skillsets is required to successfully conduct meta-analysis (Hey et al., 2009). It is important to promote multi-disciplinary teams able to contribute to the meta-analysis, or synthesis, required to solve the complex problems facing our world. Synthesis has become increasingly important for ecologists as the abundance of data and the need for the development of solutions to complex, trans-disciplinary environmental problems has grown (Marx, 2013).

It is difficult to build a team to carry out transdisciplinary synthesis activities in conventional research laboratories or institutes due to the wide range of skillsets required to accomplish any outcomes, and common to the natural resource sciences because of impediments to data-sharing (Volk et al., 2014). Analysis and synthesis of transdisciplinary data can be highly challenging, but can be enhanced by utilising data infrastructure, large-scale networks and innovative tools (Carpenter et al., 2009; Pooley et al., 2014). Synthesis builds on existing data infrastructure, research findings, and expertise obtained through the availability of technological and scientific capital to create new knowledge that is greater than the sum of the components. This is a cost-effective way of capitalising on existing data for a range of scientific problems. As a by-product, the tools and technology innovation delivered as part of synthesis activities have been used widely beyond the applications for which it was purposed (e.g. Metacat: Berkley et al., 2001; Knowledge Network for Biodiversity, KNB: Reichman, 2004; Dryad hosted by NESCent: www.nescent.org).

The maturation of ecology from individual small-scale, short-term research to a large-scale, long term, multidisciplinary field studies has given rise to the creation of synthesis centres (Reichman, 2004). One of the first ecological synthesis centres established to support trans-disciplinary synthesis was the National Centre for Ecological Analysis and Synthesis (NCEAS) in the United States of America in 1995. Since then, several synthesis centres have started in various ecosystem and biological science domains (www.synthesis-consortium.org). Synthesis centres provide a structure that brings together researchers, theorists, modellers, managers, and practitioners within a working group model to solve a common problem. The working group members use existing data to answer new questions and address complex environmental problems that require immediate solutions. The commitment of participants to share their data requires a conducive environment, flexible data policies, technical support, and appropriate data management. Synthesis centres provide the support and expertise to assist the individual and groups make the most of their own and other's data to solve complex scientific problems.

This paper will provide an analysis of the data management challenges experienced in the ecosystem science synthesis centre. This is exemplified through the Australian Centre of Ecological Analysis and Synthesis (ACEAS), established within the Terrestrial Ecosystem Research Network (TERN) to facilitate scientists, policy-makers and managers tackle some of the ecosystem questions facing Australia and the world. After analyzing the wider community's attitudes to data access, sharing, and publication at the start of ACEAS, we provide insight into the challenges of data management for synthesis, reflect on the data management literacy of participants, and discuss actions taken to mitigate challenges. It is argued that the synthesis centre functions as an incubator of data management practice.

1.1. Data management for synthesis in the ecosystem sciences

In ecosystem science, data can take quantitative or qualitative forms such as numbers, text, code, GPS co-ordinates, algorithms, models, audio, video and animations. Data management is the development and implementation of policies, plans and processes that manage these data to maintain the integrity, security and useability of data. The ideal outcome is for data to be self-described so that others can discover and re-use it effectively (Strasser et al., 2012). Effective data management is central for ecological synthesis to:

- improve efficiency and access to scientific data;
- solve complex, multi-scale environmental problems;
- allow synthesis products to be more easily accessible by a range of users; and
- enhance transparency and scientific participation in decision making (Faniel and Zimmerman, 2011); and
- enable longitudinal analyses and experiments (which require access to data collected decades ago).

As well as to enable longitudinal analyses and experiments (which require access to data collected decades ago).

Although we have unprecedented opportunities to generate new knowledge from data-intensive science, the data are not necessarily fit-for-purpose, available (open access), quality assured, and licensed properly for appropriate reuse. Changes in practices in data handling are being driven by new data formats, technological advances in hardware and software, online availability of data and increasing appreciation of value-adding as a consequence of reuse (Faniel and Zimmerman, 2011). Readiness to share data, however, differs between disciplines: for some sharing data is common whereas for others it is not (Tenopir et al., 2011; Hampton et al., 2013). Ecology has historically been a highly individual endeavour and efforts to develop a culture of data sharing have been slow (Jones et al., 2006; Hampton et al., 2013). To efficiently provide timely information requires researchers not only to continue their practice (collection of new data) but also to focus on its reusability (Hampton et al., 2013).

The dispersion, heterogeneity, and provenance of data present real technological challenges for data acquisition and use. Ecosystem science data are characteristically widely dispersed, collected from multiple sites (Reichman et al. 2011, Marx 2013) and over differing timescales. The data are often located on individual storage devices that are known only to the collecting team, and only discoverable through personal contact. Ecosystem science data are characteristically heterogeneous. Field ecologists collect biotic and abiotic information across scales, time, and space in response to the nature and behaviour of the animals and plants they study. In consequence, ecosystem data have highly variable terminologies, measurements and experimental outputs (Jones et al. 2006; Reichman et al. 2011). Tracking and recording provenance is critical to enable researchers to identify suitable data sets and enhances the transparency of scientific outcomes (Reichman et al. 2011).

One of the ways to overcome these challenges is to improve collaborative processes. Establishing trust between parties is critical, and will enhance data sharing, productivity and the generation of useful, robust outcomes (Luna-Reyes et al., 2008). A key aspect of establishing this trust is the establishment of mutually acceptable Intellectual Property Agreements (e.g. Perkmann et al., 2013; Hertzfeld et al., 2006). Once trust is established, research scientists are likely to be more motivated by opportunities for intellectual challenge, peer recognition and collaboration led by scientific peers than by financial reward or other incentives (Crane, 1969; Stern, 2004; Lam 2011). Current data management practices would be greatly enhanced by implementing a renewed focus on developing and implementing new solutions (Tenopir et al. 2011), which should include solutions based around these motivating factors of collaborative research.
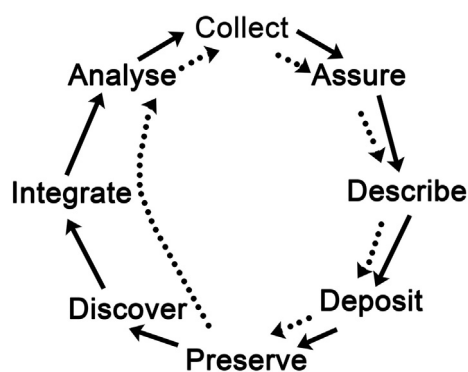
Fig. 1. The eight steps of the data life cycle from Michener et al. (2012).

1.2. Data workflow in the synthesis centre

For an individual scientist, managing one's own data is challenging, let alone discovering, manipulating and managing other's data. Participants at a synthesis centre usually concentrate on the analysis phase of the research data lifecycle (Fig. 1), and generally lack the competence or skills to publish their data for future re-use (Costello, 2009).

In ACEAS, the deposition phase involved specific attention, as one of the primary requirements of funding was that the synthesised data were published. Establishing and ensuring ownership of data are critical before analysis can proceed, and in order to deposit data, they must be described properly to be discovered and interpreted in the future. It is the task of the chosen repository to preserve the data.

For the purposes of this study the data workflow has been divided into four components, more meaningful to the analysis and synthesis process (Fig. 2):

1. Identification and acquisition (discover)[a]
2. Collation and integration
3. Analysis and synthesis
4. Publication and visualisation (deposition)[a]
   [a] Data quality assurance [throughout particularly i & iv].

The ACEAS working group activity is described against a background of the four-step data workflow (Fig. 2).

1.3. The Australian Centre for Ecological Analysis and Synthesis

The Australian Centre of Ecological Analysis and Synthesis (ACEAS) was established in 2009 to support 'disciplinary and inter-disciplinary integration, synthesis and modelling of ecosystem data to aid in the development of evidenced-based environmental management strategies and policy at regional, state and continental scales' (www.aceas.org.au). ACEAS was required, in addition, to foster trans-organisational synthesis, and the synthesised data were intended for publication. These goals presented a number of challenges, one of which was the questionable readiness of the community for such a step forward in data sharing and management. The connection with the Terrestrial Ecosystem Research Network (TERN: www.tern.org.au), a large data infrastructure observatory and repository for the ecosystem sciences, was unusual among global synthesis centres, as most have stood alone and developed their own substantial informatics teams (e.g. NESCent: Rodrigo et al., 2013). The combination of individual discipline-based researchers in relatively non-hierarchical groups, with an associated large observational infrastructure was intended to create a new paradigm in ecosystem research providing extended expertise and perspective on a range of topics.

Between November 2010 and May 2014, 42 working groups, the major focus of ACEAS activity, were supported through merit-based selection of applications from the ecosystem science and management community. These groups consisted of scientists, policy makers and managers coming together to solve challenging trans-disciplinary ecosystem problems. Projects were proposed by key investigators who nominated members according to the skills and attributes required, within the constraints of the ACEAS funding (group size, career-range, geographical, organisational representation). If successful, membership was further refined in discussion with the ACEAS Director.
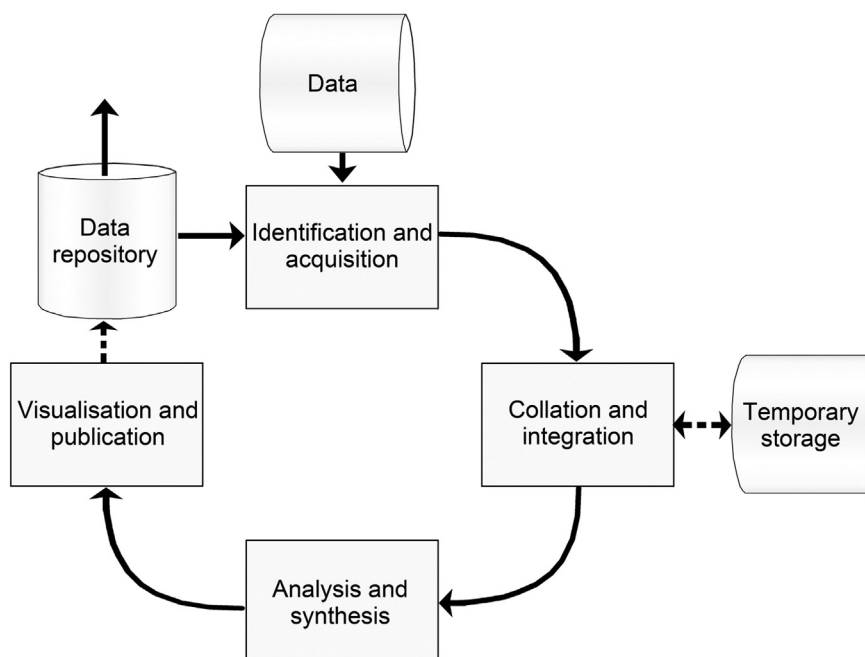

Fig. 2. The data workflow as applied in this study. Dashed arrows indicate a data deposition pathway.

The aims of ACEAS were:

- to solve or investigate ecosystem problems, which although internationally relevant, required case studies centred in Australia;
- to have meritocratic selection rather than topic-limited;
- to have a science to policy to management theme, information capture and delivery resulting in:
  - high end of inter-disciplinarily across ACEAS (remote sensing to stream water monitoring to genetics for conservation to ethno-ecosystem knowledge) and within the groups → high degree of heterogeneity,
  - high level of organisational heterogeneity within the groups.

ACEAS provided data management support through a data synthesis manager and a research assistant. Association with TERN provided additional capability through facilitated discovery of a variety of data types, expanded expertise in specialised data management, as well as a variety of data deposition and discovery options. Data sharing within ACEAS groups was based on data sharing agreements which enabled use of sensitive data to be limited to the collaboration only.

The data workflow (Fig. 2) was conducted against the background of the ACEAS working group process (Fig. 3). Interaction with ACEAS staff and with fellow group members started at application, but the face-to-face meetings were pivotal to the process. Groups generally met two times face to face for a week in a span of 12–18 months. It is important to appreciate that the ACEAS activity was itself a punctuation point in participants' research life, emerging from previous activity and feeding into the next stage of their research, and although an end-point of data delivery and publication required by ACEAS can be identified, this was, in actuality, rarely final.

A preliminary, non-cumbersome Data Management Plan (DMP) was required for the application, containing details of the datasets required, their availability, and plans for analysis and synthesis. Given the funding constraints, which limited the duration of support for groups and the number of meetings, most datasets had to be identified and available before funding was granted and indeed some funding was held over until this requirement was met. Despite this, quite often a large part of data identification occurred after the first meeting due to the crystallisation of the work through wider group discussion.

ACEAS put a lot of emphasis on the DMP's evolution during the course of the working group, not least because of an evolving identification of the data needed. The use of a DMP was designed not only to

Box 1

An example of the Data Management Plan questions for ACEAS working groups.

---

1 Acquiring data

- What data are to be used? Are they:
  in the possession of the participant
  data that need to be acquired?
  vital data that are not digital?

2 Managing data

- Which file format are the data in?
- Where will the data be stored and in what format (Ms Access, Excel workbook)?

3 Analysing data

- Do you have to combine different datasets?
- Do you need ACEAS assistance to combine your datasets?
- How will the data be processed?

4 Data to be produced

- What data will be generated in the research?
- What type of data will be created?

Metadata are developed using Ecological Markup Language (EML)

---

ensure robust data management by the group, but to assist ACEAS staff understand the requirements of the working group, identify any potential impediments in sourcing data, and allow appropriate resources to be allocated. ACEAS staff assisted in identifying and acquiring data and if appropriate technical and statistical members were not in the group, the ACEAS director would suggest members and arrange for their inclusion.

Each working group had a designated data manager who was responsible for developing and implementing their DMP. The DMP contained details of the types of data used, data standards for formatting
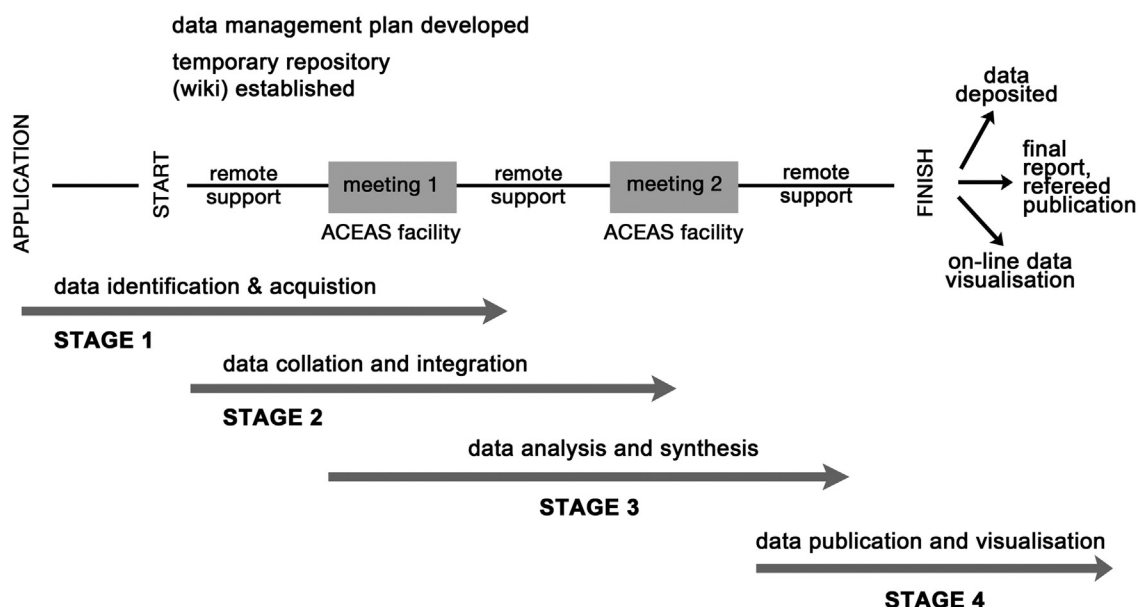

Fig. 3. Stages of the data workflow in relation to the ACEAS group experience from successful submission of application through to delivery of products.

and metadata, directives for data storage, protocols for data access and other information about the longevity of the data (Box 1). Any impediments to data use as a result of ownership were hopefully identified at this stage.

Most of the data preparation and synthesis activities happened in Stage 2 (Fig. 3). Some of the common procedures undertaken in this stage were:

(i) ensuring appropriate compliance measures for data usage,
(ii) ensuring temporary storage and access to data for all the working group members,
(iii) making the data fit to use in the synthesis activity, and
(iv) describing the data adequately so that it was understood by all the working group members.

Group members frequently brought their own data, or data from their workplace released for the purpose of the working group activity. ACEAS provided a secure, unique username and password-protected temporary repository, a wiki (www.wikispaces.com), to host these datasets accessible to all members of the group. This security was very important for many participants, as the datasets they brought were often only released for the work of the group. The wiki, through the provision of a sophisticated web content management system, enabled the stored files to be annotated, and groups could send emails and blogs from the site. It provided a one-stop-shop bulletin board for members who could not attend meetings, and an *aide memoire* for those who did. It was a valuable tool for ACEAS to communicate with the groups and promoted continuity. The uptake of the wiki as a main communication and information-sharing medium in ACEAS was a challenge, especially when working group members had their own preferences, perhaps not as good, but with which they were familiar.

Due to the heterogeneous nature of the datasets commonly used, groups needed assistance in harmonising, re-formatting and transforming datasets so they could be collated and were ready for analysis. The ACEAS Research Assistant spent a considerable amount of time in collation and integration activities.

Data analysis and synthesis occurred both during, between and after the workshops were finished. This was the stage at which the ACEAS team felt the groups would need the least informatics support, as this was their expert domain. Once the data were in the correct format, harmonised and accessible, the analysts in the group could take over.

ACEAS took Stage 4 (Figs. 2 and 3) very seriously, as the mandate of linking science outcomes to management required some innovative approaches and disciplined communication skills. ACEAS developed a sustainable data publication infrastructure so that any synthesised data was appropriately described, appropriate attribution was provided for the data creators, data were captured and published in an appropriate format to improve reusability, and make it easy for researchers to search, discover and re-use data.

All groups were required to deposit their derived, synthesised data for re-use. The Principal Investigator and/or designated data managers in the group, working closely with ACEAS staff, took responsibility for data deposition in a publically-accessible KNB-Metacat repository hosted by ACEAS with a metadata feed provided to the TERN Data Discovery Portal (TDDP) (Guru et al., 2013). Since most ACEAS working groups were ecologists, Ecological Metadata Language (EML) was used to describe the datasets. The data were shared under the open access licenses according to the TERN data licensing policy (Bradshaw et al., 2012) and assigned Digital Object Identifiers (DOIs). At time of publication, ACEAS has published 36 datasets accessible from http://aceas-data.science.uq.edu.au/knb.

Because of the ACEAS imperative of transdisciplinarity (*sensu* Rice, 2013; Polk, 2014), to ensure all stakeholders were informed, reports for each meeting and a final report, or précis of the group's activities and achievements (including a heading, 'How will this affect ecosystem science and management') were also required. To make the data delivered more interpretable, where relevant, ACEAS also undertook to support groups to visualise the data delivered. As many of the data products had spatial contextual information, this lent itself to publishing them on a Web Mapping Service (WMS), with the map feature published through a Web Feature Service (WFS) accessible through http://aceas-data.science.uq.edu.au/portal, http://mammalviz.tern.org.au/, and the Indigenous Biocultural Knowledge working group's web delivery http://www.aibk.info. This latter web site was established to deliver data for the Australian Indigenous community so they could take over management of the site once the working group activity was finished. The data itself are downloadable in a range of relevant file formats (SHP, GML, CSV, PDF). This method of delivery showcased the synthesis outcomes and promoted a culture of data sharing. At closure, twelve such web visualisations were produced. An example of this delivery is shown in Fig. 4.

The data workflow is fundamental to the management of ecosystem data. The ability of the researcher to appreciate, understand and participate in all stages of the data workflow is uncertain, and in this paper the role of ACEAS in improving the awareness and ability of members of the ecosystem science and management community in data management is examined. This paper reports on the insights that both participants and staff gained from their ACEAS experience that inform our understanding of the attitudes, readiness and behaviour of the community at large.

2. Methods

The methodology used in this study fell into three main components: (i) an initial survey of the community at large (the potential users of the synthesis centre): 'Community Survey', (ii) a 'working group analysis' drawn from selected synthesis centre groups illustrative of each phase of the data workflow, and (iii) a survey of the ACEAS community's experience of the challenges and learning about data management from their working group experience: 'ACEAS community survey'. Also at our disposal were data from routine surveys administered to participants after workshops, and reflective journals kept by staff who used them to inform practice and enhance their understanding (Ortlipp, 2008): 'routine feedback'.

2.1. Community survey

As a background to the synthesis centre and data infrastructure initiative, a survey was conducted to assess the perceptions and practice of inter- and trans-disciplinary research collaboration in the Australian ecosystem science and management community, and their attitudes to and practice of data access, sharing and publication. This was used as a baseline to better understand the community with which ACEAS was working. The questions were inspired by the work of Tenopir et al. (2011). Data were collected from the ecosystem science and management community via a semi-qualitative questionnaire about perceptions of research collaboration, recent collaborative activity and attitudes to data sharing (Table 1). These were accompanied by situational questions (discipline, organisation, qualifications, gender and so on).

In November 2011, hard copies of the surveys were distributed to all 650 delegates at the 2011 Ecological Society of Australia conference. Subsequently, a stratified snowball sampling method was used to reach approximately 3200 community members between December 2011 and February 2012. The precise response rate was not determinable.

Overall, 751 respondents answered at least one question, representing a response rate of around 20%. Of these, 75 identified themselves as ACEAS participants. At that time eight groups and one sabbatical fellow had had meetings (a total of 133 participants), and none had finished their work, so it is considered they remained reflective of the wider community. The data were analysed using SPSS with bootstrapping applied to modify the non-normal data.
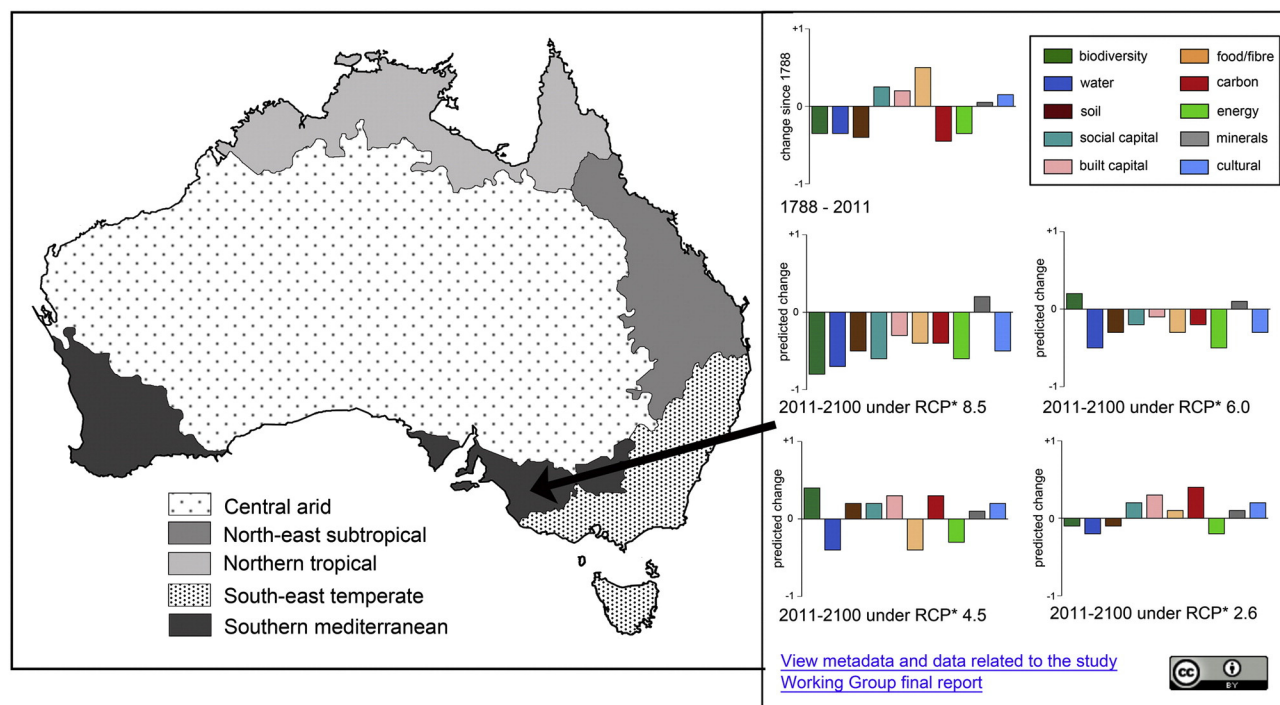
Fig. 4. A modified display of a characteristic on-line data visualisation from one of the ACEAS working groups, showing the interactive feature and interpreted data delivery with links to the raw, synthesised data and metadata description. Each sector in the map has a unique interpretation, linked to a data product. Other links employed include specific geospatial reference points, in one case more than 700 across Australasia (http://aceas-data.science.uq.edu.au/portal).

2.2. Working group analysis

A session was organised at a symposium for all ACEAS working groups in May 2014 in which selected ACEAS working groups presented their experiences against each of the steps of the data workflow. Using a method of purposive sampling, we invited representatives of working groups who, from our observations, illustrated a particular step well. Two groups, as Case Studies (CS), were selected for each step, each of whom have delivered or are intending to deliver data products.

Data identification and acquisition: CS1. Seagrass mapping and climate change <http://www.aceas.org.au/index.php?option=com_content&view=article&id=97&Itemid=98>; CS2. Adaptation pathways for aquatic plants <http://www.aceas.org.au/index.php?option=com_content&view=article&id=93&Itemid=95>.

Data collation and integration: CS3. Animal telemetry <http://www.aceas.org.au/index.php?option=com_content&view=article&id=101&Itemid=103>; CS4. Northern quolls <http://www.aceas.org.au/index.php?option=com_content&view=article&id=129&Itemid=137>.

Data analysis and synthesis: CS5. Impacts of climate change on endemics <http://www.aceas.org.au/index.php?option=com_content&view=article&id=98&Itemid=152>; CS6. Drought-associated dieback <http://www.aceas.org.au/index.php?option=com_content&view=article&id=102&Itemid=104>.

Data publication and visualisation: CS7. Aerobiology and allergic responses <http://www.aceas.org.au/index.php?option=com_content&view=article&id=113&Itemid=114#top>; CS8. Indigenous biocultural knowledge <http://www.aceas.org.au/index.

Table 1
The questions posed around data in the 2011–12 survey of the ecosystem science and management community. All questions required a response on a 5 point scale, with 1 being strongly disagree to 5 strongly agree.

Broad question field	Specific questions
Data use (question 9, Tenopir et al., 2011 survey).	Q1 Lack of access to data generated by other researchers or institutions is a major impediment to progress in science.
	Q2 Lack of access to data generated by other researchers or institutions has restricted my ability to answer scientific questions.
	Q3 Data may be misinterpreted due to poor quality of the data.
	Q4 Data may be used in other ways than intended.
Data sharing (question 10, Tenopir et al., 2011 survey, 2 sub-questions re: use of a data repository omitted).	Q5 I would use other researchers' datasets if their datasets were easily accessible.
	Q6 I would be more likely to make my data available if I could place conditions on access.
	Q7 I would make my data available if any users of the data would cite it, in the same way as a paper I produce would be cited.
	Q8 I am satisfied with my ability to integrate data from disparate sources to address research questions.
	Q9 I would be willing to share data across a broad group of researchers who use data in different ways.
	Q10 It is important that my data are cited when used by other researchers.
	Q11 It is appropriate to create new datasets from shared data.

Table 2
The questions posed in the 2014 survey of selected representatives of the ACEAS community.

Question	Response type
Q1 What was your experience in data management before the ACEAS working group?	Pick list
Q2 Did your group(s) use the ACEAS Data Management Plan?	Pick list and open-ended comments
Q3 How did your group manage your primary data with respect to storage, sharing and data ownership?	Open-ended comments
Q4 From your experience, what was the most challenging part of the data management cycle for your Working Group?	Pick list and open-ended comments
Q5 What part of the data management workflow did you learn most about through your Working Group experience?	Pick list and open-ended comments
Q6 Did you or your group submit a data outcome to ACEAS?	Pick list and open-ended comments
Q7 In the future, do you intend to have a DMP (Data Management Plan) for your next project even if it is not mandatory?	Pick list and open-ended comments

php?option=com_content&view=article&id=103&Itemid=105#top>.

One of the authors (Treloar) performed the role of rapporteur during this session and the presentations were recorded and interrogated afterwards. Through this method key challenges and experiences were identified and reported as case studies.

2.3. ACEAS community survey

A survey was designed to enquire about use and attitudes to formal data management, acquisition and management of the data, what were felt to be the most difficult parts of the data workflow and what respondents had learnt most from in their synthesis centre experience (Table 2). The survey was administered on-line and invitations with a link to the survey instrument emailed to 70 members of the ACEAS working group community: each Principal Investigator and the group data manager (if identified) or one to two co-investigators with known activity in data management. All but two responses were received after 10 days, and closed after 14 days.

The data from this survey were collated in Excel and quantitative data were analysed using SPSS with bootstrapping applied to modify the non-normal data. The qualitative responses were analysed using grounded theory, with key themes identified and reported.

2.4. Routine feedback

At the end of each ACEAS workshop, formal feedback was systematically sought to assess participant satisfaction, group dynamics and performance, attitudes to data sharing and collaboration, expected

products and outcomes, and ways to improve ACEAS function. Only a core group of questions were asked throughout the survey period, including "How do you feel your association with ACEAS has differed from other workplace collaborations".

Specific questions for data sharing and collaboration were developed for the latter phase of the survey period (based on a period of exploratory questions):

(i) Do you think you were able to synthesise existing data better through the ACEAS process than you would have otherwise?
(ii) How do you feel your attitude towards collaboration has changed as a result of this workshop?
(iii) How do you feel your attitude towards data sharing has changed as a result of this workshop? and
(iv) Do you think the synthesis undertaken by ACEAS groups promotes a culture of synthesis, collaboration and data sharing?

The data from these sources were collated in Excel and quantitative data were analysed using SPSS with bootstrapping applied to modify the non-normal data. Qualitative responses were analysed using grounded theory, with key themes identified and reported.

3. Results

3.1. Community survey

Seven hundred responses were obtained overall for the data questions. Not all replied to every question. Respondents were mainly environmental scientists (61%) and ecologists from academic institutions (mainly universities), and had a higher degree (Table 3).

Not all of these categories were collected for ACEAS participants (e.g. age or qualifications), but the profile was reflective of the larger community (Table 4).

Responses were positive to most questions (Fig. 5). The most positive response was that data citation should be a requirement for data sharing (Question 10: Fig. 5). Feelings about the relevance of citation was also asked in Question 7, and responses to both were broken down according to: (i) employment category and (ii) gender. Academics considered it most important that their data were cited when used, while those from non-government organisations were least concerned (academics scoring 4.43 and NGOs 4.02 with all other groups between: F = 6.19, p = 0.000, Tukey's post hoc test at alpha ≤0.05). Females considered they would be more likely to make their data available if users were to cite it (Question 7) than males (scoring 4.12 versus 3.93 on a 5 point scale: F = 7.60, p = 0.006, Tukey's post hoc test at alpha ≤0.05).

Five hundred and fifty two respondents gave their gender and 236 were female and 316 male. A significantly greater proportion of males had a higher degree than females (70%:56%, Pearson's chi squared p = 0.000) while gender balance was equal for the bachelor's degree, females felt slightly, but significantly more strongly than males, that access to data was a major impediment to progress in science (Question 1:

Table 3
Profile of respondents to the community survey. Numbers and the percentage of the surveyed community are shown (content within rows is not correlated).

Age (years)	Frequency (%)	Employment[a]	Frequency (%)	Qualifications[a]	Frequency (%)	Discipline[a]	Frequency (%)
18–27	39 (7)	Academic	243 (43.7)	School or trade certificate	16 (2.1)	Biology	79 (15.4)
28–32	97 (17.5)	Government	147 (26.4)	Bachelor's degree	180 (24.0)	Environmental management	75 (14.6)
33–40	137 (24.7)	Non-government organisation inter alia	46 (8.3)	Higher degree	352 (46.9)	Environmental science and ecology	316 (61)
41–50	143 (25.8)	Private company	83 (14.9)			Other	42 (8.2)
51–75	139 (25)	Research organisation	37 (6.7)				

[a] Academic' includes postgraduate students; 'non-government organisation' includes Natural Resource Management agencies and statutory authorities; 'private company' includes mining industry employees inter alia. 'Bachelor's degree' includes those with honours or a postgraduate certificate. Discipline was self-nominated. 'Other' includes social scientists, biometricians and physicists.
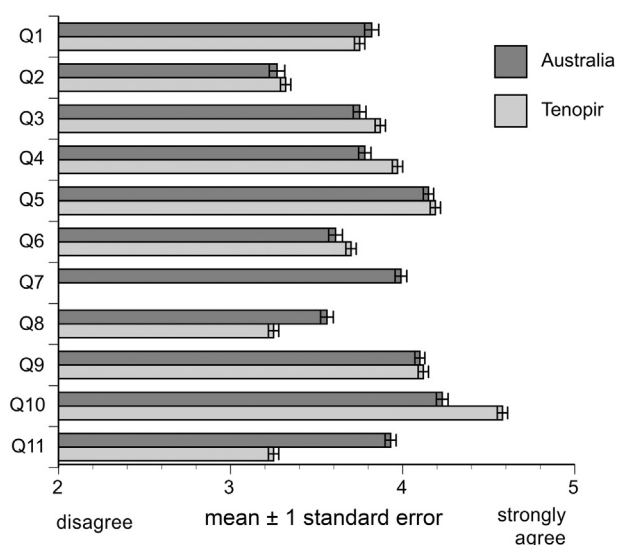
Fig. 5. Responses to the eleven questions posed to the Australian ecosystem science and management community about data use and data sharing (see Table 1 for text of questions). These are plotted alongside the responses to the questions reported in Tenopir et al. (2011). Responses were on a Likert scale of 1–5 (strongly disagree to strongly agree).

3.92 versus 3.69 on a 5 point scale: F = 6.61, p = 0.01, Tukey's post hoc test at alpha ≤0.05) although not for their own research (Question 2).

Respondents were somewhat concerned about their ability to use data due to poor data quality (Question 3, 3.75 on the 5 point scale) and were a little concerned about their ability to integrate data from diverse sources to answer their research questions (Question 8, 3.56 on the 5 point scale). Creating new datasets from blended data was positively regarded overall (Question 11, 3.93 on the 5 point scale).

Comparisons with the results of Tenopir et al. (2011) showed that the Australian community was, at the time of the surveys, more ready to create new from existing data (Question 11), and less concerned about having their data cited (Question 10) than the predominantly United States of America cohort. The proportion of academics in the Australian survey was half that of Tenopir et al. (2011) (43.7% to 80.5%) perhaps associated with the higher score in the USA survey, as academics were more inclined, as shown in the Australian survey, to consider citation important than other workplace categories. The Australians were also more confident of their ability to manipulate data (Question 8) but there was consistency in the remaining responses (Fig. 5).

3.2. Working group analysis

In this section, we report the experiences of the representative working groups against each of the different stages of the data workflow.

Table 4
Profile of ACEAS participants. Numbers and their percentage of the survey community are shown.

Employment[a]	Frequency (%)	Discipline[a]	Frequency (%)
Academic	408 (54.6)	Biology	27 (11.7)
Government	152 (20.4)	Environmental management	43 (18.6)
Non-government organisation *inter alia*	43 (5.8)	Environmental science and ecology	148 (64.1)
Private company	29 (3.9)	Other	13 (5.6)
Research organisation	115 (15.4)		

[a] All categories include an international component. The sample size for discipline was 231. This question was not asked until the latter part of ACEAS' routine surveys.

3.2.1. Identification and acquisition

The Principal Investigator of an ACEAS working group identified both the data and expertise to help answer an environmental question.

Significant challenges with the identification of suitable data were found to be:

- lack of metadata and hence ability to assess the suitability of the data; and
- limited access to data.

Significant challenges with the acquisition of data were found to be:

- permission to obtain the data;
- time and financial constraints; and
- quality assurance.

Working groups tried to overcome these challenges by employing various strategies (Table 5).

3.2.2. Collation and integration

Once working groups identified and acquired their data, it had to be brought together and transformed into a suitable format so that it was fit for purpose and analyses could be conducted. The two selected groups reported similar difficulties (Table 6).

3.2.3. Analysis and synthesis

When the significant challenges of the first two steps of the data workflow are overcome, ACEAS working groups can begin to analyse and synthesise the data. These collaborative activities are often breaking new ground due to the amount of data required for analyses, and the high degree of multi-disciplinarily. This required creative, adaptive solutions (Table 7).

3.2.4. Publication and visualisation

ACEAS working groups are expected to publish their synthesised datasets publicly so that it may be re-used by other scientists to tackle other research questions. The ACEAS Data Portal provides a means to visualise the produced datasets in a way that is more conducive to uptake by scientists, managers, policy makers and the general public. Publication, however, presented several challenges (Table 8). A minority of ACEAS groups reached this stage.

In summary, there were several challenges based on the observations made by the sampled working groups. These included:

- a difficulty to find suitable data with required spatial and temporal resolution,
- lack of metadata with additional contextual information,
- limited data available in open access that fitted the project requirements,
- difficulty to trace data provenance to get additional information,
- reluctance to share primary data because of fears of mis-use, inappropriate attribution, need to release intellectual property, need to publish before sharing etc.,
- in certain cases lack of domain standards to describe and share data, and
- difficulties of data discovery in open access repositories.

These difficulties impeded the ability of the groups to conduct synthesis: in particular the unknown quality of much of the existing data led to high uncertainty.

3.3. ACEAS community survey

Thirty-one responses were received for this survey, a response rate of 44%. Each respondent (R) was given a unique reference number

Table 5
Responses from two groups selected due to their experiences in data identification and acquisition.

Case study 1 — seagrass mapping and climate change (Kilminster et al., 2015)	Case study 2 — adaptation pathways for aquatic plants
Datasets were brought together to develop a nation-wide, spatially explicit risk assessment for seagrass habitat. Steps included: (i) to develop a habitat map, (ii) to identify risks, (iii) to acquire suitable risk layers, (iv) assign risk, and (v) run the assessment. Sources of data included government, university and non-government organisations. Limitations included: • time constraints in obtaining permission to use data, • different temporal and spatial scales between datasets, • lack of metadata — need for corporate knowledge, and • lack of clear and uniform variables. As a result of the deficiencies in some of the data, the group developed proxy datasets, illustrating the multi-disciplinary problem-solving skills utilised in synthesis.	This group encountered problems identifying and acquiring highly dispersed data across the full continental range required. Appropriate datasets were primarily identified through working group contacts, facilitated by the collaborative model. These were primarily intensive, small-scale, dispersed, data collection activities together with larger-scale environmental monitoring data. Combining these dispersed data provided the capacity to infer what is happening over a range of wetland environments across the continent. Effects of climate change required access to hydrological climate change data, sometimes difficult to access in areas not already modelled. At individual level, government employees are committed to long-term data (e.g. Driver et al. 2013) but at whole-of-government not always possible. Key data workflow activities included ensuring that the data were consistent with national metadata standards, ensuring taxonomic synonymy, and acquiring climatic variables of the right spatial and temporal scale. It is not always possible to identify and acquire a suitable dataset but there is an opportunity to reach out to various communities as a response so that in the future such data collection activities could be undertaken to obtain such data.

Table 6
Responses from two groups selected due to their experiences in data collation and integration.

Case study 3 — animal telemetry (Campbell et al., 2015)	Case study 4 — Northern Quoll
This group quickly realised that the data challenges to achieve their research outcomes were considerable, and they determined they could make a major contribution to the field by publishing an interactive spatial presentation of the available literature. This meta-analysis was the first step towards open access to animal telemetry data to enhance collaboration across the research community. The group found that: • researchers were reluctant to share their data due to concerns about re-use, • re-use of data was challenging due to the question-specific context of the original studies, and • metadata was inadequate in many cases to be confident in re-use. In order to increase the data available for their analysis, the group provided researchers, in exchange for their data, a suite of specialised tools for spatial analysis.	This group's basic task before analysis could proceed was to establish a database of population and environmental parameters of a rare and spatially distributed species, the Northern Quoll (*Dasyurus hallucatus*). The main barriers to this purpose were: • obtaining data of suitable scale and across the full geographic representation, • data mismatches; • different research purposes in data collection; and • different data collection methods. Integrating these datasets to allow for adequate statistical analysis of population was a highly challenging endeavour because some data could not be modified to suit the modelling input requirements. As an example of the effort required to prepare data, a supplementary meeting was organised, as greater than 50% of workshop time in their first meeting was consumed with data blending and preparation for analyses. Notably, data preparation extended beyond meeting one and was still partly required at the second meeting. Despite this, the collated datasets produced a significant amount of data for over 4500 individuals from 15 locations across northern Australia, and was hugely successful in setting up a baseline for data collation and integration and getting the data fit for ongoing and new research purposes.

Table 7
Responses from two groups selected due to their experiences in data analysis and synthesis.

Case study 5 — impacts of climate change on endemics	Case study 6 — drought-associated dieback
Spatially and temporally consistent, standardised species diversity and demography data were required for over 220 animal and plant species (the largest collation of genetic datasets on terrestrial vertebrates and plants). Significant challenges were encountered due to a lack of: • adequate metadata [for DNA sequences and spatial co-ordinates], • difficulties obtaining information from the authors to acquire missing metadata, • formatting data correctly for analysis • had to establish consistent, standardised analysis techniques, and • having the technical skills to write code. But for the first time there is an overview of how Australian terrestrial vertebrates and plants responded to past climate events.	This group intended to determine the relationship between plant water relations, drought tolerance traits and climate through the analysis and synthesis of existing data. Barriers encountered included the lack of a comprehensive database of tree mortality events, and data that were available tended to be patchy, incomplete and not suitable for integrated analysis and synthesis. This group, comprising a wide range of expertise, developed an approach using statistical probability that could deal with the data workflow challenges and broaden the understanding of drought effects on forest ecosystems as well as improve future data management practices.

Table 8
Responses from two groups selected due to their experiences in data publication and visualisation.

Case study 7 — aerobiology and allergic responses (Davies et al., 2015)	Case study 8 — indigenous biocultural knowledge (Pert et al., 2015)
There were many challenges with publication and visualisation of synthesised data with this group, which was highly transorganisational and transdisciplinary in nature. There was strong reluctance to publically release primary data from a technological and cultural viewpoint:	This group was aware of several interesting challenges to data publication at the start of their collaboration, particularly related to the nature of 'data' and data ownership. Indigenous knowledge is largely unwritten and unrecorded, requiring strong collaborations and a hands-on approach to its discovery. Indigenous peoples control, maintain and protect their traditional knowledge, and their intellectual property over such knowledge.
• many journals won't publish articles if the data has already been released for open access, • priority of publication and the fear of empowering a competitor or commercial partner, and • insecurity about the quality of data.	The group therefore targeted existing knowledge that was already publically available but difficult to locate, with a view to presenting it in a way that resonated with the indigenous and non-indigenous communities. The result was a spatial database of existing publically available indigenous bio-cultural knowledge that can be viewed by anyone interested in identifying this knowledge (www.aibk.info)
This group had to overcome impediments related to ownership of synthesised data before publication, as a result of:	
• difficulties in contacting a data contributor to obtain their permission for synthesised data release, • unknown ownership of a contributing dataset, • lack of institutional data publication policies; • data licensing restrictions, and • confusion with respect to correct attribution of the original data providers.	

and this is quoted as 'R1, 2, 3' etc. Of these, the majority of respondents had experience as data managers, either moderate (26%) or they 'did it all the time' (42%). Twenty six percent said they had no or little experience. Despite this strong representation of experienced data managers, 35% of respondents reported they did not prepare a Data Management Plan for the group (despite the template being available both in their workshop packs and on the wiki as a separate page: see Box 1) and 35% said they did not know whether it had been done or not.

Of those whose groups used a Data Management Plan (29%) it was stated to be most useful for:

• identifying data requirements: "Useful guidelines to what was needed in the meta-data sections. Helped clarify confusion as to what was needed for the data" (R6) and "very useful. Critical for consistency and repeatability of sites in different bioregions" (R5),
• organisation: "It helped me organise the structure of the datasets and work out how to manage their storage, quality and use" (R1), "As a framework for the diverse group to keep us on track" (R15); and
• determining storage requirements: "It assisted us to store and discover data between the two workshops when some members were different." (R13).

One respondent said although they had prepared a DMP it was not useful "… because of the wide range of data sources accessed, each with different ownership requirements. It was also difficult to determine what constituted data from an ACEAS perspective — raw data or some kind of product" (R31).

Of those whose groups did not prepare a DMP, only a few provided comments, but of those that did, the type of work they were doing was a reason not to: "We weren't collating or managing large amounts of data that we needed to synthesise, or which would be directly useful to others in the future" (R31).

Although ACEAS provided ample and secure interim storage of primary data for analysis and synthesis on a family of wikis, groups predominantly stored their datasets externally (79%), on laptops, in Dropbox, on work servers, with all of which they were familiar, but which proved on occasion to be less than optimal when the custodian dropped out. One respondent reported they made hard copies of their data for security.

When it came to sharing data, Dropbox was the most favoured mode (34%) with the wikis provided for the group (28%) second, with emails, ftp servers and discussion also mentioned. Some mentioned they would share data after they were cleaned and compiled, and there was some mention of the security of data within the group and licensing concerns ("only some datasets shared between owners and analysts", R17).

Data ownership would seem to be one of the most critical matters that can be solved through preparation of a Data Management Plan. Resolution of ownership concerns was assisted by ACEAS ("we used the ACEAS guidelines, and all participants must agree about all uses and publications" R28) and on one occasion consultation with university legal officers was mentioned. The scale of access reported ranged from open access (50%) with clear metadata descriptions of provenance, to continuing to reside with the original owner (non-synthesised data: 27%), to restricted access (23%: "Negotiated with each provider …" R6).

3.3.1. Challenges

The most challenging component of the data workflow for the groups was the identification and acquisition of data followed by data
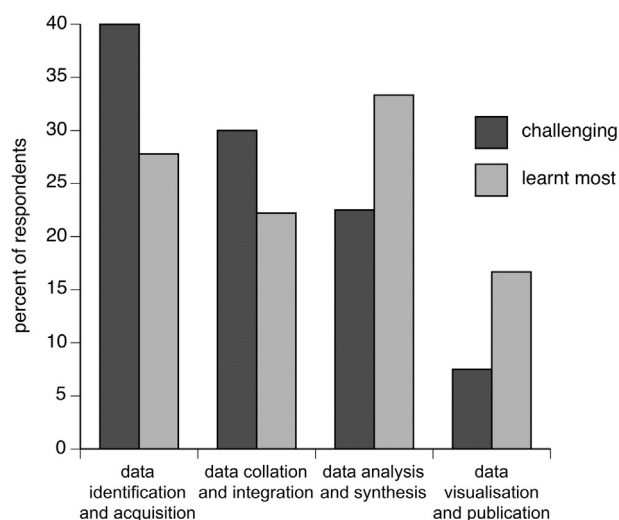
Fig. 6. The most challenging components of the data workflow, and the component in which respondents felt they learnt the most.

collation and integration (Fig. 6). "Although most members were happy to share data, obtaining and collating data from other sources were difficult" (R25), and "data ownership restrictions were challenging" (R6). The scale of the challenge depends on how novel the experience was, and "In our field there is not much support for data-sharing" (R13).

The quality of the data for the purpose was a major stumbling block, with data needing standardisation and transformation before it could be used: "We were primarily addressing questions for which there was very little primary data. Agreeing on the appropriateness of disparate data sources was a challenge" (R31). "With so much data of different coverages (extent), quality, purpose, resolution and methods under which datasets were developed, it is difficult to identify the right for your purpose. In some cases, is not well documented (metadata) or accessible. For some of our work [the] data were just not available" (R20). "We had data from various sources and formats that had to be cleaned and integrated [before use]" (R10).

Sometimes the difficulties can mean data may not be pursued or fully analysed in the time available "It takes a lot of time to wrangle diverse data into congruent formats, this is very difficult and time consuming, no one has much spare time" (R2), "data were in different formats and as they were plot-based, the different plot sizes presented challenges to what could be done with the data." (R6), "… a huge amount of work in standardising the data. We did not get so far as to analyse the data during the workshops, but will in the future" (R25). This was a point where ACEAS staff could assist "Took time with ACEAS staff to sort out as it was not straight forward" (R12). Indeed time was a critical feature in 11% of responses to this question "Not sure any part is challenging, just requires the effort to do it" (R19).

The use of legacy data provided challenges "Legacy data on outmoded formats and need to bring up to date re: taxonomic and formatting standards" (R24); "On occasions physically accessing data on hand written records and obsolete hardware was additionally challenging" (R1). Even when data were acquired successfully, they sometimes needed additional treatment to analysed "Required custom scripts due to 4D data volume" (R16).

Having a diverse group membership did prove useful when trying to pull data together (and this was a step that had to be somewhat advanced before the groups were funded), as "Trying to collate multiple datasets at similar spatial scale resolution across a large area (Australia) was challenging, not always obvious where data were housed, representatives from different agencies and states enabled identification" (R23). Difficulties in identification were followed by difficulties in access "Some datasets were difficult to locate, some were also difficult to get permissions to use and synthesise." (R21).

3.3.2. Acquired knowledge

Participants learnt the most about analysis and synthesis, followed by data identification and acquisition (Fig. 6). Comments by 20 of the 31 respondents also noted new knowledge around data manipulation, data acquisition and data publication. Analysis and synthesis were something several respondents were most unfamiliar with: "I had done least of this previously" (R28), and "I had never worked on such a synthesis project before, so it was all a new experience for me" (R30).

There were several comments about new skills learnt through the working group process "… through this process we learned different ways to analyse existing data." (R20); "Learnt new statistical approaches" (R29); "Learned a lot about extracting information directly from group knowledge (i.e. expert elicitation)" (R3); and "I honed my skills in liaising and negotiation with primary data holders" (R1). The strength of some of the tools brought to bear was commented on: "I've become a strong supporter of quantitative literature searches based on my experience in two ACEAS working groups where I think the approach has yielded very strong data and allowed firm statements to be made that otherwise could be more easily challenged." (R31).

The ability to work in a diverse group was a source of new knowledge "Spatial analysis and visualisation are not my area of expertise so
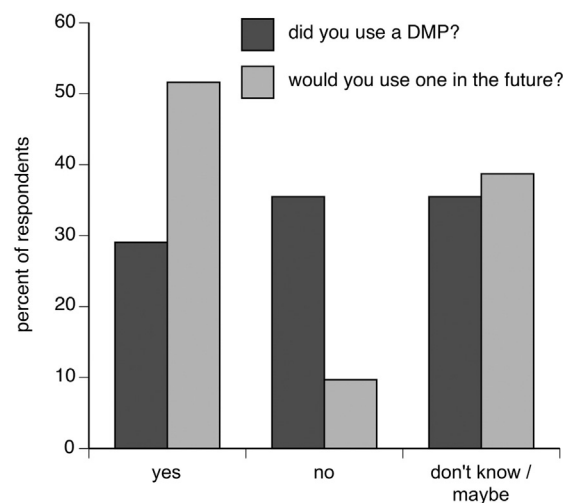

Fig. 7. Actual use of a Data Management Plan for the ACEAS group (ACEAS WG: don't know) and after the experience, intended future use of a DMP (in the future: maybe).

it was great to have a multi-disciplinary group that could each contribute different strengths and skills". (R15) "Through the work with the other workshop participants, I found very useful to see other people's data information systems and the way they organised and integrated different datasets." (R20).

The scale of work possible in a multi-disciplinary group appeared to be a step above participants' past experience "we've been working though ways of integrating data at the national scale. This is a bigger and more unruly task than I'd ever attempted previously" (R9), and "It was a challenging data analysis project that called on my past acquired skills but to be used in a different environment and applied to a different project." (R10).

Data publication and visualisation were a new experience for many, "the importance of metadata, and making the data open access via publication" (R19), and they appreciated ACEAS assistance "Working with the TERN-ACEAS team and the TERN Data Discovery Portal helped us acquire the spatial data we needed for our study and produce the online spatial outputs" (R22), and "Learnt via ACEAS staff about the challenges of integrating data into a Portal" (R12).

We anticipated that one of the most novel components of the workflow would be data delivery and visualisation which 45% of the respondents groups had done through their ACEAS activity, and this involved a specific type of ACEAS support which initially was not in place and the process was "slow" (R25). Respondents whose groups had uploaded data reported "ACEAS sent a data manager to work with us — so the experience was relatively good" (R6). "The process was involved and protracted primarily because of access issues. The ACEAS staff were helpful, skillful and attentive in facilitating the formatting and uploading process. They were responsive to the specific situation of our working group" (R1). "What we submitted was very brief, and not digested much, just broad outlines. However ACEAS was very helpful in visualising what we did submit" (R25), with the time being limiting in product delivery "Worked very well. Once the system and workflow was established I became the bottleneck" (R5).

When asked whether, after the experience of working with ACEAS, they would use a Data Management Plan in the future 52% said they would, a greater proportion that had done at the start of their interaction (Fig. 7). The undecided (don't know whether my group prepared a DMP or maybe we might in the future) remained the same percentage throughout. Reasons given by those who said they would use a DMP in future were predominantly for conflict resolution (ownership, storage, promoting transparency) and for its value in project planning and organisation. Three respondents mentioned the increasing mandatory need for a DMP either via funding agency or for publication "Empirical pooled

data are easily managed but the products of these syntheses probably would be easier professionally if a DMP was in place" (R4).

For those who answered 'maybe' the use of a DMP appeared superfluous "We seemed to do this implicitly. We discussed the aims and how we would use the data to reach those aims" (R10), or conditional "The basics should apply to any exercise, but for some simple exercise or with small groups the plan may be more a statement of intent or process, rather than anything more complex or detailed. Hence, it may not appear to be a plan, depending on how people view plans or planning" (R21).

Three respondents (10%) said they would not use a DMP in the future, and one commented that existing protocols were sufficient "I take responsibility for storage of data I generate within my institution, and I consider this is the level at which data should be managed" (R29).

3.4. Routine surveys

Three hundred and thirteen responses, a response rate of 76%, were received to the question "How do you feel your association with ACEAS has differed from other workplace collaborations". Each respondent (RS) was given a unique reference number which is quoted as 'RS1, 2, 3' etc.

Enhanced collaboration and networking provided by ACEAS were most commonly remarked upon (32%). Many felt a spirit of collaboration was more evident in the ACEAS groups than in other workplace collaborations: for example, "Clear emphasis on collaboration and synthesis and support to focus on that task" (RS360); "ACEAS has definitely actively, and successfully, worked to contribute to collaboration process, much more than might usually have happened by the workshop conveners/funding organisations" (RS144) and "The ACEAS workshop was a very targeted and structured process for encouraging a large group to collaborate. Getting this level of input in a short space of time often doesn't work very well in many workplaces" (RS174).

The design of the working groups to encompass a wide range of disciplines and organisational representation was also commented on as a distinguishing factor, with comment such as "There was a wider geographical and skills/professional range: it's good not to only talk to scientists in your town, but include managers and people from other regions" (RS232); "The distribution of skills allowed roadblocks to be quickly overcome and a wide range of outcomes to be completed" (R456); and "Bringing together wide group of people to work on common purpose with a product at the end. Multi-disciplinary was a real plus." (RS392). These approaches enhanced the type of product "Improved the reasons and aims of collaboration. Actually working on the problem during [the] meeting enhanced my desire to collaborate. Broken down barriers between propeller heads [sic] and users." (RS269).

Thirty percent of respondents to the routine surveys commented on the uniqueness of the support that ACEAS offered, some of which freed them to 'get on with it': "Excellent organisation and support enables participants and particularly the organisers to focus on the workshop content." RS109; and "Excellent organisation so we can focus on working all day. Clear objectives" (RS324). A substantial number explicitly commented on the difference in the support provided to enable better analysis and synthesis "The level of support to focus on synthesis, which is usually a last minute activity at the end of a project/programme" (RS360); "ACEAS wasn't just funding the WS but encouraged to collaborate in a productive way (wiki, support, etc)" (RS263); and "Much more contact with ACEAS staff; other collaborations tend to be self-contained." (RS373). Vision was expressed by RS359 "[the synthesis centre can] remove/support tedious tasks like bulk data management".

Statements that specifically mentioned improved data analysis and synthesis and data sharing practices as a result of the ACEAS experience were less abundant (8%), with comments such as "It was purely about coming together to work on a common problem and share information. This would have been harder without ACEAS involvement" (RS57),

illustrating the close link between supporting collaboration and data (information) sharing. "Working on datasets in the room allows immediate discussion with experts on issues that would otherwise be put aside." (RS7); and "Focus on synthesizing information at a high level. Most workshops I have attended focus on a specific issues, which may be only relevant for discrete locations/or scientific problems." (RS170). The ACEAS experience was noted to "Allow for very rapid generation of data" (RS154).

There was a small minority (2%: seven participants) who said they discerned no difference in the way ACEAS conducted its business to other workplace collaborations. They did not elaborate. A further 2% said that the model reflected their experiences at NCEAS, the Australian Research Council's Australia–New Zealand Vegetation Function Network or with one of Australia's Cooperative Research Networks[1] (Coastal). "ACEAS has taken on the mantle of the successful ARC-vegetation workshops and developed them." (RS283); "Strongly analogous to NCEAS, which is hardly surprising!" (RS108); and "It is actually much the same as the coastal CRC — good" (RS118).

In the latter phase of the routine surveys, specific questions were asked to flesh out the difference between the ACEAS experience and other collaboration and data synthesis activities. Sixty nine responses were received, a response rate of 98.5%. On a scale of 1–10 where 1 is 'strongly less likely/never or not at all' and 10 is 'strongly more likely/extremely so', responses for all questions were positive (Fig. 8). The strongest response was a near maximum score (9.07 ± 0.002) for the development of a culture of synthesis, collaboration and data sharing through the ACEAS experience. Most felt their attitude to data sharing had changed as a result of their ACEAS experience (at 7.09 ± 0.002), and more felt that their ability to synthesise data was much better than they would have been able to otherwise achieve (8.46 ± 0.001).

4. Discussion

The community in which ACEAS was working was not dissimilar in its data collaboration profile to that surveyed by Tenopir et al. (2011) (Fig. 5), despite a significantly lower proportion of academics in the Australian cohort. Academics across the globe are subject to performance criteria highly dependent on citation rates, and this was reflected in the responses to the Tenopir et al. (2011) survey, and was a particular concern for the academic cohort within the Australian respondents (Section 3.1). Improved data sharing seemed to be conditional on adequate protection for the data supplier. The survey showed that Australian participants were more ready to create new datasets from shared data than those from the predominantly United States of America cohort: in other words they were receptive to the synthesis centre concept (Fig. 5). Concern about poor access to data was expressed by the stakeholder community (Fig. 5), reinforced by many comments from the ACEAS community (Section 3.2).

A lack of lived experience in data practice is borne out through the results of the targeted surveys of the ACEAS participants (Sections 3.2 and 3.3), as ACEAS participants report their challenges (Fig. 6 and Sections 3.2.1 and 3.3.1) and intended change in behaviour in response to personal exposure to the benefits of a collaborative data workflow (Fig. 7). This extended across all groups, as the consistent theme of inadequate metadata for re-use was mentioned in all case studies.

The major challenge for the majority of the working groups was sourcing primary data (case studies 1 and 2; Fig. 6). This was one of the reasons that ACEAS required groups to supply as much of this information as possible at application. As mentioned in Section 1.3, some groups were rejected due to their low likelihood of acquiring the required data in time. The diversity of the ACEAS working groups improved chances of direct data sourcing and encouraged accurate descriptions of metadata on collection and use. ACEAS was able to

---

[1] http://www.business.gov.au/grants-and-assistance/Collaboration/CRC/about-the-program/Pages/default.aspx.
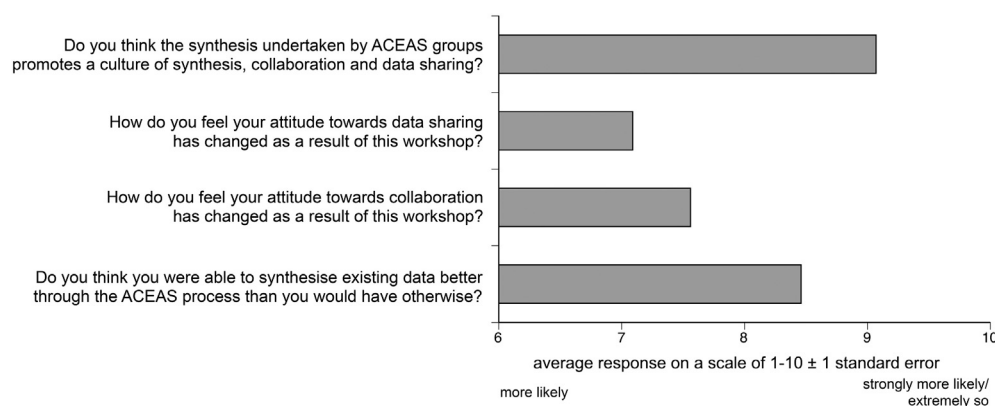
Fig. 8. Attitudes to data management and collaborative practices reported through routine surveys of ACEAS Working Group participants. Standard errors are too small to be visible at this scale.

help groups acquire data through its good relationship with state and federal government agencies, leveraging access to the Terrestrial Ecosystem Research Network's repositories, other data repositories, and sometimes by paying data retrieval fees to expedite access.

The time and effort required to prepare data for synthesis were often underestimated, particularly when dealing with data only previously used by the data owner and which had limited metadata and idiosyncratic labelling and descriptions (e.g. comments in Table 6). Working groups, with the assistance of ACEAS staff, spent considerable time transforming, re-formatting, transferring data to different drives, checking data quality, and even digitising data. This experience was arguably good, however, as these future data publishers learnt first-hand what was involved to allow their data to be easily re-used. This might encourage them to return to their domain colleagues and establish accepted minimum elements of metadata that will make datasets more reusable.

Once the data were ready, data sharing occurred in a range of ways largely dependent on the participants' past experience, from Dropbox and the supplied wikis, to emailing datasets to one another (Section 3.3). One factor overlooked by participants was that if data were accessible by ACEAS personnel on a shared but secure repository it was able to be curated properly (stored, labelled, annotated), and recovered easily in case of participant failure. Such failure occurred in a couple of instances, and centralised hosting of temporary data enabled ACEAS to step in and facilitate the other members of the group pick up the work.

One of the major impediments to data sharing is trust, which predisposes data sharing. Often group members were selected because the Principal Investigator thought they would have suitable data because of their publications, but did not know them personally prior to the meeting. Indeed, most of the working group members met face-to-face for the first time during the first workshop, and although this was good for network-building (a major positive outcome reported in the routine feedback surveys), it meant time was needed to develop the trust required for confident data sharing. Building trust was easier in domains such as molecular biology where sharing and analysing data in teams and across labs are common than in other, more insular disciplines such as forest eco-physiology. Synergies between such groups occurred when they met concurrently. This provided insights into another domain and its practices, breaking down hard barriers with beneficial results. The ACEAS initiative provided a unique opportunity as commented by a respondent in the routine surveys taken after every meeting (R424) "I feel it is rare for people in science to share data and I think it has been great to achieve individually but can't always. Pulling together of many disciplines is rare."

The experience of data delivery and analysis and synthesis were reportedly the least challenging stages of the data workflow, but ones in which the participants learnt the most (Fig. 6). Despite a generally collaborative spirit, transdisciplinary collaboration for analysis and synthesis was novel for many. Many participants commented on the benefits of working with an inter-disciplinary team of scientists and managers, which shed new light on how their work could be used to fit the purpose of another, promoting innovation and extending the usefulness and applicability of their own data and research activities. In one instance it was the first time that dedicated field ecologists fully appreciated the requirements of statisticians, with their specific expectations of data quality, type, and the sample size required for analyses. In several cases where hard data were lacking, expert elicitation, modelling, and Bayesian methods were used to finalise the outcomes.

Particular mention of the support of ACEAS staff was made at data delivery, as the process of delivery was not well known (Fig. 6), and there was reluctance to release data if it was not of the highest quality and was cleared through sometimes complicated ownership hurdles (Table 8). Visualisation was expected to be technically the most novel component of the workflow for participants, and support was anticipated for this phase. The use of spatial delivery was intuitive for the groups, as often it was modelled on what was published in their papers (e.g. Murphy et al., 2013). Correct attribution of source data for release proved a great stumbling block for a number of groups (e.g. case studies 7 and 8, Table 8). The technical aspects of preparation of data for re-use, namely disciplined organisation of files and creation of associated metadata, somewhat unexpectedly required close guidance from the ACEAS staff before the data could be released. The closer the participant's experience of the need for good metadata in the acquisition and blending of diverse data for analysis (stage 2), the more willing they were to spend the time on this stage. This is a learned skill, and with time, the support required for this stage should be reduced. It was important, however, that ACEAS provided quality assurance, ensuring that metadata was accurate and understandable (requiring close collaboration between the expert, the informatics team and the ACEAS Director as publisher), that all the appropriate releases were in place, and for the visualisations, that they were, within our limited means, of suitable standard that they were interpretable and usable by the community.

The preparation of a Data Management Plan (DMP) would have alleviated many of these challenges, but there was some resistance, with 70% reporting their group had either not done a DMP, or it was so irrelevant to them they did not know if one had been done. Many of the 42% of respondents who said they managed data all the time must not have used a DMP to do so. One of the most difficult feats for the ACEAS staff was promoting the development and implementation of a DMP. Appreciation of the value of the DMP was often greatest at data delivery stage where attribution and ownership of data became important, and when the metadata were being written for publication (e.g. case study 7). If the DMP had been properly prepared at an earlier stage, data delivery would have been much easier. Respondents to the

survey showed they appreciated this, with a 20% increase in determination to use a DMP in the future as a result of their ACEAS experience (Fig. 7).

Tension due to the greater part of the work being accomplished outside the meetings (Fig. 3) was recognised by many groups, and efforts were made by both them and ACEAS to mitigate this effect through support for remote collaboration through the wikis, additional meeting support and staff contact. Face to face collaboration was critical, however, allowing time to build the vital relationships, establish communication pathways and data management practices, share analysis methodologies, discuss publications, determine data products, and prioritise future efforts. The process of building these relationships targeted core motivating factors for scientists within group activities; peer interaction and recognition within a common discipline of interest, intellectual challenge and enhanced publication opportunities (see introduction). This approach sustained the groups between and beyond the meetings and has resulted in the many and varied products.

As was clear from the routine feedback surveys (Section 3.4), the support provided by the synthesis centre allowed domain-specialists to concentrate on their area of expertise and to learn and collaborate with their colleagues in tackling their specific transdisciplinary science to management question. Transdisciplinarity requires disciplinary masters to be informed by other discipline areas to create a new understanding, and adding sophisticated data management skills onto this already demanding environment would dilute the core pursuit if the support provided by a centre was not present. This does not mean to say that data expertise should not be acquired, and is not a valuable attribute for any scientist or manager, but it is the marriage of the many disciplines and expertise, including data, that makes for innovative and novel outcomes.

5. Conclusion

To tackle the substantial environmental problems facing us today, researchers need to be adept in quantitative and qualitative transdisciplinary synthesis activities, working with environmental managers and policy makers. This is challenging, and does not occur spontaneously.

ACEAS working groups identified many issues that impeded their synthesis work. The culture of the ecosystem science and management community needs to be changed so standard practice and metadata standards are followed when collecting and describing data, and appropriate tools and technology are applied to data handling, including its publication. Data re-purposing was recognised as risky because of the uncertain quality of much data and hence unsuitability for synthesis. Adopting good practice will improve traceability of the manipulations made to the raw data and hence improve re-usability. It is important to engage with the research community to educate them of advantages of data sharing while acknowledging their apprehensions.

A weakness in the ecosystem sciences is the broad range of data collection protocols and a lack of disciplined data description (metadata), making capitalising on existing data and knowledge difficult if not impossible. Synthesis centres offer a unique opportunity to test hypotheses using incubation funding. Through using existing data, new knowledge can be generated without reliance on establishing new experiments or infrastructure. ACEAS has offered an opportunity for transdisciplinary synthesis projects in ecosystem science to flourish, with around 750 people from around the world participating in 42 ACEAS working groups, delivering 36 datasets and 12 data visualisations. The legacy of the intervention includes an improved culture of and skills in collaboration, synthesis and data sharing. This has carried forward to twelve nationally competitive grants emerging directly from ACEAS work, around twenty multi-authored papers (at time of publication) and greatly enhanced networks across boundaries not otherwise spanned.

Synthesis centres provide a training ground for data re-use of which data management is a key component, something that many do not think they need, but when it comes to practice a subject in which

domain specialists (at least in this cohort) are demonstrably ill-experienced. The experience of participation in a synthesis centre has clearly improved all parties' understanding of, and skill in, data management, and its importance in enabling the delivery of products. The support offered makes the acquisition of new knowledge, the discipline of data management and delivery of products meaningful and more effective than when working in isolation or with a small group of colleagues reinforcing existing practices.

References

Berkley, C., Jones, M.B., Bojilova, J., Higgins, D., 2001. Metacat: a schema-independent XML database system. Proc. 13th Int. Conf. Sci. Stat. Database Manag.. IEEE Comput. Soc.

Bradshaw, A., Deed, J., Dickinson, D., Guru, S.M., Phinn, S., Clancy, T., 2012. Data licensing: the Terrestrial Ecosystem Research Network (TERN) approach. Presented at the eResearch Australasia Conference, 2012, Sydney.

Campbell, H.A., Beyer, H.L., Dennis, T.E., Dwyer, R.G., Forester, J.D., Fukuda, Y., Lynch, C., Hindell, M.A., Menke, N., Morales, J.M., Richardson, C., Rodgers, E., Taylor, G., Watts, M.E., Wescott, D.A., 2015. Finding our way: on the sharing and reuse of animal telemetry data. Sci. Total Environ. (this issue).

Carpenter, S., Armbrust, E.V., Arzberger, P., Chapin, F.S., Elser, I., Hackett, J., Ives, A., Kareiva, P., Leibold, M., Lundberg, P., Mangel, M., Merchant, N., Murdoch, W.W., Palmer, M.A., Peters, D., Pickett, S.T.A., Smith, K.K., Wall, D.H., Zimmerman, A.S., 2009. Accelerate synthesis in ecology and environmental sciences. Bioscience 59 (8), 699–701.

Costello, M.J., 2009. Motivating online publication of data. Bioscience 59 (5). http://dx.doi.org/10.1525/bio.2009.59.5.9.

Crane, D., 1969. Social structure in a group of scientists: a test of the "invisible college" hypothesis. Am. Sociol. Rev. 34 (3), 335–352.

Davies, J., Beggs, P., Medek, D., Erbas, B., Thibaudon, M., Katelaris, C.H., Haberle, S., Newbigin, E., Newham, R.M., Huete, A.R., 2015. Trans-disciplinary synthesis of the impact of pollen aerobiology on health in Australasia. Sci. Total Environ. (this issue).

Driver, P., Raine, A., Foster, N., Williams, S., 2013. Ecological monitoring to support Water Sharing Plan evaluation, and protect wetlands of inland New South Wales, Australia. Ecol. Manag. Restor. 14 (3), 187–193.

Ellison, A.M., 2010. Repeatability and transparency in ecological research. Ecology 91 (9), 2536–2539.

Faniel, I.M., Zimmerman, A., 2011. Beyond the data deluge: a research agenda for large-scale data sharing and reuse. Int. J. Digit. Curation 6 (1), 58–69.

Guru, S., Shen, X., Love, C., Treloar, A., Phinn, S., Wilkinson, R., Brady, C., Isaac, P., Clancy, T., 2013. Sharing Australia's nationally significant terrestrial ecosystem data: a collaboration between TERN and ANDS. IEEE 9th International conference on eScience, Beijing, China, pp. 53–60 http://dx.doi.org/10.1109/eScience. 2013.28.

Hampton, S., Strasser, C., Tewksbury, J., Gram, W., Budden, A., Batcheller, A., Duke, C., Porter, J., 2013. Big data and the future of ecology. Front. Ecol. Environ. 11 (3), 156–162. http://dx.doi.org/10.1890/120103.

Hertzfeld, H.R., Link, A.N., Vonortas, N.S., 2006. Intellectual property protection mechanisms in research partnerships. Res. Policy 35 (6), 825–838.

Hey, T., Tansley, S., Tolle, K. (Eds.), 2009. The Fourth Paradigm. Microsoft Research.

Jones, M.B., Schildhauer, M.P., Reichman, O.J., Bowers, S., 2006. The new bioinformatics: integrating ecological data from the gene to the biosphere. Annu. Rev. Ecol. Evol. Syst. 37 (1), 519–544.

Kilminster, K., McMahon, K., Waycott, M., Kendrick, G.A., Scanes, P., McKenzie, L., O'Brien, K., Lyons, M., Ferguson, A., Maxwell, P., Glasby, T., Udy, J., 2015. Unravelling complexity in seagrass systems for management: Australia as a microcosm. Sci. Total Environ. (this issue).

Lam, A., 2011. What motivates academic scientists to engage in research commercialization: 'gold', 'ribbon' or 'puzzle'? Res. Policy 40 (10), 1354–1368.

Luna-Reyes, L.F., Black, L.J., Cresswell, A.M., Pardo, T.A., 2008. Knowledge sharing and trust in collaborative requirements analysis. Syst. Dyn. Rev. 24 (3), 265–297.

Marx, V., 2013. Biology: the big challenges of big data. Nature 498 (7453), 255–260.

Michener, W.K., Allard, S., Budden, A., Cook, R.B., Douglass, K., Frame, M., Kelling, S., Koskela, R., Tenopir, C., Vieglais, D.A., 2012. Participatory design of DataONE—enabling cyberinfrastructure for the biological and environmental sciences. Ecol. Inform. 11, 5–15.

Murphy, B.P., Bradstock, R.A., Boer, M.M., Carter, J., Cary, G.J., Cochrane, M.A., Fensham, R.J., Russell-Smith, J., Williamson, G.J., Bowman, D.M.J.S., 2013. Fire regimes of Australia, a pyrogeographic model system. J. Biogeogr. 40 (6), 1048–1058. http://dx.doi.org/10.1111/jbi.12065.

Ortlipp, M., 2008. Keeping and using reflective journals in the qualitative research process. Qual. Rep. 13 (4), 695–705.

Perkmann, M., Tartari, V., McKelvey, M., Autio, E., Broström, A., D'Este, P., Fini, R., Geunae, A., Grimaldi, R., Hughes, A., Krabelh, S., Kitsong, M., Llerenai, P., Lissonij, F., Salter, A., Sobrero, M., 2013. Academic engagement and commercialisation: a review of the literature on university–industry relations. Res. Policy 42 (2), 423–442.

Pert, P., Ens, E.J., Clarke, P.A., Locke, J., Packer, J.M., Turpin, G., 2015. An online spatial database of Australian Indigenous Biocultural Knowledge for contemporary natural and cultural resource management. Sci. Total Environ. (this issue).

Peters, D.P.C., 2010. Accessible ecology: synthesis of the long, deep, and broad. Trends Ecol. Evol. 25 (10), 592–601.

Polk, M., 2014. Achieving the promise of transdisciplinarity: a critical exploration of the relationship between transdisciplinary research and societal problem solving. Sustain. Sci. http://dx.doi.org/10.1007/s11625-014-0247-7.

Pooley, S.P., Mendelsohn, J.A., Milner-Gulland, E.J., 2014. Hunting down the chimera of multiple disciplinarity in conservation science. Conserv. Biol. 28 (1), 22–32.

Reichman, O.J., 2004. NCEAS: promoting creative collaborations. PLoS Biol. 2, 0311–0312.

Reichman, O.J., Jones, M.B., Schildhauer, M.P., 2011. Challenges and opportunities of open data in ecology. Science 331 (6018), 703–705.

Rodrigo, A., Alberts, S., Cranston, K., Kingsolver, J., Lapp, H., McClain, C., Smith, R., Vision, T., Weintraub, J., Wiegmann, B., 2013. Science incubators: synthesis centers and their role in the research ecosystem. PLOS Biol 11 (1), e1001468. http://dx.doi.org/10.1371/journal.pbio.1001468.

Rice, M., 2013. Spanning disciplinary, sectoral and international boundaries: a sea change towards transdisciplinary global environmental change research? Curr. Opin. Environ. Sustain. 5, 409–419.

Stern, S., 2004. Do scientists pay to be scientists? Manag. Sci. 50 (6), 835–853.

Strasser, C., Cook, R., Michener, W., Budden, A., 2012. Primer on Data Management: What You Always Wanted to Know. A DataONE, publication http://dx.doi.org/10.5060/D2251G48.

Tenopir, C., Allard, S., Douglass, K., Aydinoglu, A.U., Wu, L., Read, E., Manoff, M., Frame, M., 2011. Data sharing by scientists: practices and perceptions. PLoS One 6 (6), e21101. http://dx.doi.org/10.1371/journal.pone.0021101.

Volk, C.J., Lucero, Y., Barnas, K., 2014. Why is data sharing in collaborative natural resource efforts so hard and what can we do to improve it? Environ. Manag. 53, 883–893.